How Trends Are Affecting Databases - Panel Discussion

Betty Perloff, USDA-ARS-HNIS

The U.S. Department of Agriculture (USDA) has developed and continually manages several data bases containing information about foods. This information covers food names; food descriptors; food formulations and recipes; the composition of foods; food yields; nutrient retentions for different types of foods; factors for deriving energy, protein, and fatty acid values of foods; and weights for various measures of foods.

Food data bases are accessed by various USDA systems, each of which, in turn, generates its own unique food-related data sets. For example, the National Nutrient Data Bank system uses all of the different types of food information listed above during its processing to produce reference data sets on the composition of foods. Another system, Survey Net, uses food names, descriptions, recipes, and food weights for coding dietary intake data from nationwide food surveys; links the intake data to food composition information; and produces data sets on the consumption of both foods and nutrients. A third system, the Food Grouping System, utilizes the data sets produced by the Nutrient Data Bank and Survey Net to report food and nutrient consumption data after the foods have been grouped by various food characteristics. The ingredients in mixtures are taken into account during the food grouping process.

Several factors affect the priorities and current direction of data base activities at HNIS. First, the shear numbers of foods and the increasing diversity of new foods are having a tremendous impact on the amount of resources needed for management of food data bases. Second, the need to process food consumption survey data quickly requires intensive efforts to obtain information about new foods in a very short period. Third, the need to respond to food safety concerns requires knowing not only the nutritional value of foods but also the specific ingredients and processing steps associated with foods.

In addition, all of these requirements must be balanced against the need to ensure that the data bases are as accurate as possible. In October of last year, the General Accounting Office issued recommendations for improving USDA's food composition data. While full implementation of their recommendations is desirable, the funding that would be required for additional food analyses by the government far exceeds the available budget. In the meantime, USDA must meet the demands for information about foods by carefully setting priorities for expending the limited public funds available for food analyses and by developing techniques to provide adequate estimates about the composition of foods that cannot be analyzed.

The data base most often affected by the numbers of new foods and the diversity of those foods is the food coding data base used within Survey Net. Foods reported by survey respondents in the Continuing Survey of Food Intakes by Individuals are matched against the foods in this data base, and foods reported in the National Health and Nutrition Examination Survey III (NHANES III) are also linked to the codes which represent the foods. Each year several thousand new foods enter the market. We hear about many of these foods for the first time when their names or descriptions are reported by respondents in the food consumption survey. When a new food is reported, staff immediately begin locating information about it, such as what ingredients are in it, what its composition is, and how much various portion sizes weigh.

The National Center for Health Statistics provides information about new foods appearing on NHANES III. This coding data base contains over 7,000 individual items with their own unique food codes, and over 5,000 additional food names or descriptions are linked to those codes. It also contains over 28,000 weights of portion sizes and common household measures associated with the foods in the data base.

One of the most time-consuming aspects of processing dietary intake data from nationwide food surveys is dealing with uncodable items, which are usually new foods or new combinations of existing foods. One of the main goals of developing Survey Net was to facilitate this aspect of processing survey data. Survey Net operates at both the survey contractor's site and at USDA. Coded intake records are sent electronically to USDA. In turn, updated food data bases (new food descriptions, weights, and recipes) are sent back to the contractor.

When a new item is encountered for the first time, it is written to a special file of uncodable items, where it automatically receives its own code number. Once a new food has been listed on this special file, it is available for retrieval and selection during the coding of subsequent food records. Each time an item is selected from this file, its unique code number is recorded on the intake record instead of a food code. The file of uncodable items is transmitted to HNIS along with the intake records, and the Survey Net team decides if each item needs a new code or if an existing code can be used.

Another feature of Survey Net is recipe modification. During food coding, the recipe for an existing food can be modified to record more specific information when supplied by the respondent. One of the main uses of this feature is to record the specific types of fat and milk used in recipes.

The recipe modification feature also facilitates collecting the detailed information needed for dealing with food safety concerns. For example, the Environmental Protection Agency needs better estimates of water consumption. Existing recipes for reconstituted items, such as infant formulas, can be modified when the dilution does not follow the package direction.

Because of the large numbers of new foods reported on the survey and because nutrient values are not available for all nutrients in all foods, considerable effort is required to estimate nutrient values, and HNIS has been focusing on automating more of its estimating procedures. Many values are derived through calculations using recipes and estimated formulations. Staff have discussed at previous Data Bank Conferences the technique for estimating formulations using ingredient lists and partial nutrient profiles. The Nutrient Data Research Branch has been writing other special computer programs to derive values based on existing data for similar foods by making adjustments in moisture, fat, or other specified components. These programs will also be used to assist in tracking the derivation of values. As part of the planning process for the redesign of the Nutrient Data Bank system, staff have developed an extensive set of data derivation codes that will be used for indicating the types of processes used to generate values.

We have also begun tracking some types of changes for items in the nutrient files used with our Survey Nutrient Data Base in Survey Net. When new values are inserted, we indicate if they represent data improvements or actual changes in foods. When real changes occur in foods, such as changes in fortification levels of nutrients in breakfast cereals, we retain the older values in the data base with the appropriate dates attached to the values.

In the near future we believe a great deal of our data base maintenance will be in response to changes taking place because of new labeling requirements. For example, new descriptors required for certain foods will require changes in the food coding data base, and new fortification levels expected for some

breakfast cereals will affect the nutrient data base. For the longer range, we believe calculations of mixtures using formulations will continue to be an important part of our process. All of our calculation processes will eventually be formalized within the National Nutrient Data Bank system and codes indicating the types of calculation processes will be included with the final values. We are hopeful that recent attention focused on food composition data will result in more emphasis being given to better analytical methods, including methods that are less expensive, so that eventually the quantity and quality of food composition data that are truely needed can be realized.